CFS CANADIAN FOREST SERVICE
SCF SERVICE CANADIEN DES FORÊTS
cfs-scf.nrcan-rncan.gc.ca

Using conditional spatial randomization to identify insect infestation hot spots

by Trisalyn Nelson[1] and Barry Boots[2]

Mountain Pine Beetle Initiative Working Paper 2007–12

Natural Resources Canada, Canadian Forest Service,
Pacific Forestry Centre, 506 West Burnside Road, Victoria, BC V8Z 1M5
(250) 363-0600 • www.pfc.cfs.nrcan.gc.ca

Natural Resources Canada | Ressources naturelles Canada

Canada

Canada

Using conditional spatial randomization to identify insect infestation hot spots

by Trisalyn Nelson[1] and Barry Boots[2]

Mountain Pine Beetle Initiative Working Paper 2007–12

[1.] Department of Geography, University of Victoria
Victoria, British Columbia, V8W 3P5, Canada (email: tnelson@mail.geog.uvic.ca)

[2.] Department of Geography and Environmental Studies, Wilfrid Laurier University
Waterloo, Ontario N2L 3C5, Canada (e-mail: bboots@wlu.ca)

Mountain Pine Beetle Initiative PO #3.05

Previously published in the *Journal of Geographical Systems*

Natural Resources Canada
Canadian Forest Service
Pacific Forestry Centre
506 West Burnside Road
Victoria, British Columbia V8Z 1M5
Canada

©Her Majesty the Queen in Right of Canada 2007
Printed in Canada
2007

Library and Archives Canada Cataloguing in Publication

Nelson, Trisalyn, 1975-
Identifying insect infestation hot spots : an approach using conditional spatial randomization/ by Trisalyn Nelson and Barry Boots.

(Mountain Pine Beetle Initiative working paper ; 2007-12)
Includes bibliographical references.
ISBN 978-0-662-45372-7
Cat. no.: Fo143-3/2007-12E

1. Pine--Diseases and pests--British Columbia--Houston Region--Statistics. 2. Mountain pine beetle--British Columbia--Houston Region. 3. Pine--Diseases and pests--British Columbia--Houston Region--Mathematical models. I. Boots, Barry II. Pacific Forestry Centre III. Title. IV. Series.

SB945.M78N44 2007 634.9'7516768 C2007-980069-6

Abstract

Epidemic populations of mountain pine beetle highlight the need to understand landscape-scale spatial patterns of infestation. Observed infestation patterns were explored using a randomization procedure conditioned on the probability of forest risk to beetle attack. Four randomization algorithms reflecting different representations of the data and beetle processes were investigated. Local test statistics computed from raster representations of surfaces of kernel density estimates of infestation intensity were used to identify locations where infestation values were significantly higher than expected by chance (hot spots). Investigation of landscape characteristics associated with hot spots suggests factors that may contribute to high observed infestations.

Key words: Conditional spatial randomization, kernel density estimation, mountain pine beetle, local statistics

Résumé

L'ampleur épidémique des populations de dendroctone du pin ponderosa souligne l'importance de comprendre la répartition spatiale des infestations à l'échelle du paysage. La distribution observée des infestations a été examinée en utilisant une méthode de randomisation conditionnée par la probabilité que la forêt soit attaquée par le dendroctone. On a étudié quatre algorithmes stochastiques reflétant différentes représentations des données et des processus d'infestation par le dendroctone. Pour déterminer les endroits où les valeurs relatives à l'infestation étaient beaucoup plus élevées que celles auxquelles on s'attendait par la méthode aléatoire (points sensibles), on a utilisé des statistiques d'essais locaux. Ces dernières ont été calculées à partir de représentations matricielles des superficies où l'on avait effectué des estimations de densité par la méthode du noyau en vue d'obtenir l'intensité de l'infestation. L'examen des caractéristiques du paysage associées aux points sensibles propose des facteurs qui pourraient contribuer aux fortes infestations observées.

Mots-clés: Randomisation spatiale conditionnelle, estimation de densité par la méthode du noyau, dendroctone du pin ponderosa, statistiques locales

Contents

1 **Introduction** 1
2 **Study area and data** 2
3 Forest Risk Model 3
3.1 Stand Susceptibility 3
3.2 Beetle Pressure 5
3.3 Forest Risk 6
4 **Identifying and exploring hot spots** 6
4.1 Randomization Algorithms 8
4.2 Test Statistic 9
4.3 Exploring hot spots 10
5 **Results and discussion** 11
5.1 Conditional Randomization 11
5.2 Comparisons with landscape characteristics 12
6 **Conclusions** 15
Acknowledgements 16
Contacts 16
References 17

Tables

Table 1. Age factors used in the Shore and Safranyik forest risk model (1992). 4
Table 2. Stand density factors used in the Shore and Safranyik forest risk model (1992). 5
Table 3. Location factors used in the Shore and Safranyik forest risk model (1992). 5
Table 4. Diameter at breast height (dbh) factors used in the Shore and Safranyik forest risk model (1992). 5
Table 5. Grid-based mountain pine beetle infestation size for use in the Shore and Safranyik forest risk model (1992). 6
Table 6. Beetle pressure calculations for the Shore and Safranyik forest risk model (1992). 6
Table 7. Overview of conditional spatial randomization algorithms used in the study. 8
Table 8. The percentage of pixels with risk greater than 0 that were hot spots in Morice and its sub-areas, from 1996 to 2002. 11

Figures

Figure 1. Location of the Morice Timber Supply Area showing sub-areas used in the study. 2
Figure 2. An overview of the procedure for identifying statistically significant cells (hot spots). 7
Figure 3. The probability of forest risk calculated for the Morice Timber Supply Area for 1995 and 2001. 10
Figure 4. Number of conditional randomizations less than observed value. Black indicates statistically significant (α=0.01) locations (hot spots) 12
Figure 5. Distribution of forest age by sub-area in Morice Timber Supply Area: hot spots = solid line; all locations with pine in sub-area = dotted line. 13
Figure 6. Distribution of percent pine by sub-area in the Morice Timber Supply Area: hot spots = solid line; all locations with pine in sub-area = dotted line. 13
Figure 7. Distribution of elevation by sub-area in the Morice Timber Supply Area: hot spots = solid line; all locations with pine in sub-area = dotted line. 14
Figure 8. Distribution of aspect in the South sub-area of the Morice Timber Supply Area: all locations with pine and hot spots by year. 14

1 Introduction

The mountain pine beetle (*Dendroctonus ponderosae* Hopkins) is a species native to western North America and is important for maintaining healthy pine forests. By infesting and killing lodgepole pine (*Pinus contorta* var. *Latifolia*), mountain pine beetles naturally disturb the forest and aid succession (Safranyik et al. 1974; Parminter 1998). However periodic epidemics, resulting from warmer temperatures and an abundance of mature pine, can have significant impacts on environment, economy and communities. Currently, the largest recorded mountain pine beetle epidemic in British Columbia is underway, with more than 7 million ha of lodgepole pine being affected (Westfall 2004). This situation highlights need for a landscape-level understanding of spatial patterns in mountain pine beetle infestations. Currently, despite considerable work at a stand scale (e.g., Mitchell and Preisler 1991; Barclay et al. 1998), there is a lack of research relating to this topic. Thus, we propose that exploratory analysis be undertaken as an appropriate initial step.

A common strategy for exploratory analysis of such spatial patterns is to evaluate characteristics of an empirical pattern relative to those expected from the realization of a null/hypothesized model (Getis and Boots 1978; O'Sullivan and Unwin 2003). As the properties of the model are known, this procedure may be used diagnostically to explore characteristics of an empirical pattern. More formally, one or more test statistics may be computed from the empirical pattern and the probability of these occurring under the null model may be evaluated. In exploratory spatial pattern analysis, the most frequently used null model is that of complete spatial randomness (Boots and Getis 1988).

One means of evaluating a test statistic is to use a permutation test, in which the test statistic is calculated for the original data and then the observed data are permuted and the test statistic computed for each permutation (Edgington 1995; Conner and Simberloff 1986). Collectively, the empirical test statistic and those generated from the permutations provide a reference distribution for evaluating the empirical test statistic. The probability (or significance) of the empirical test statistic occurring under the null model is equal to the proportion of permutations that have test statistic values greater than or equal to the empirical test statistic. When the data are permuted randomly, we refer to this as a randomization test (Fortin and Jacquez 2000).

For spatial data, the null model of complete spatial randomness is equivalent to complete randomization in which the data values are assigned at random to locations within the study region. That is, the data values are assumed to be located uniformly and independently over the entire study region. However, there are situations in which complete randomization is inappropriate, as either or both of the assumptions of uniformity and independence are likely to be violated (e.g., Legendre and Fortin 1989; Stine and Hunsaker 2001). For example, the probability of events occurring is unlikely to be constant over an environmentally inhomogeneous study area. This is the case for mountain pine beetle infestations, reflecting variations in forest conditions across a landscape (Safranyik et al. 1974). In such circumstances, using complete randomization would typically increase the number of Type I errors associated with the test statistic (Legendre and Fortin 1989; Legendre et al. 2002). To overcome this, complete randomization may be conditioned using *a priori* knowledge of the study region or the events under investigation.

There are several ways to condition a spatial randomization. Of relevance here are those that use *a priori* knowledge of spatial variation in the probability of the occurrence of the phenomenon being investigated. This approach has been used most extensively in epidemiology, where randomizations are conditioned on the population at risk (Besag and Newell 1991; Kulldorff et al. 2003). Similar examples are found also

in ecology, where the statistical significance of the spatial pattern of an animal is tested using randomizations restricted by habitat (Davis et al. 2000).

For spatial patterns, the test statistic may be one that summarizes the entire data set (global) or one that is computed for every data site (local; Fotheringham and Brunsdon 1999). Local statistics are particularly useful for identifying differences—rather than similarities—in spatial data via mappable measures (Fotheringham 1997; Boots 2002). Using local methods, we can undertake exploratory, diagnostic investigations of spatial variation in probabilities of the empirical test statistic.

In this paper, our goal is to explore patterns of mountain pine beetle infestations using a conditional randomization procedure that recognizes that the probability of infestation occurrence is not spatially uniform. A stand-scale, mountain pine beetle model of forest risk (see Section 3) was used to derive the probabilities. We represent both the empirical and the randomized data in the form of raster surfaces generated using kernel density estimators. Our local test statistics are the observed intensity values for each pixel in the raster surface. Thus, we can evaluate the likelihood of occurrence of the observed value of the test statistic for each pixel relative to the values for the corresponding pixels in surfaces created from the randomized data. In this way, we can identify pixels (locations) with empirical values that are significantly smaller and larger than those expected by chance. However, in this paper, we consider only the latter as these are of much more interest from a management perspective. We refer to such locations as hot spots. By investigating characteristics of the landscape at these locations, we are able to suggest factors that may contribute to high observed infestations.

2 Study area and data

Figure 1. Location of the Morice Timber Supply Area showing sub-areas used in the study.

One area impacted by the current infestation in British Columbia is the Morice Timber Supply Area (henceforward referred to as Morice; Figure 1). Covering an area of approximately 1.5 million ha, Morice is dominated by lodgepole pine and spruces (*Picea*). In Morice, mountain pine beetle infestations have been monitored since 1995 with point-based, global positioning system (GPS) helicopter aerial surveys. These surveys use indicators of pine mortality—mainly changes in crown foliage colour—to monitor mountain pine beetle activity. Once attacked, trees change from green to yellow, to red and, eventually, to gray (Safranyik et al. 1974). In this way, clusters of visually infested trees are identified and estimates of the number of infested trees they contain are made. Cluster centres are mapped as points using a GPS. The maximum area represented by a point is 0.031 km squared, equivalent to a circle with a radius of 100 m. From 1995 to 2002, a total of 43,751 data points were identified during aerial surveys.

Onset of infestation is not contemporaneous throughout Morice. Whereas much of the northern part of Morice was already infested in 1995, infestations did not occur in most of the southern part until 2000. Anticipating that this might affect our analyses, we divided the district into three sub-areas on the basis of initial infestation dates (See Figure 1). Unless the results of our analyses are consistent throughout Morice, they are presented separately for the north, middle, and south sub-areas (see Section 5).

Two supplementary data sets were used in this study: inventory data from the British Columbia Ministry of Forests, and a digital elevation model (DEM). Forest inventory data are primarily generated via aerial photo (1:15,000) interpretation and were last updated in the mid-1990s. (For details of the forest inventory procedure, see Leckie and Gillis 1995). Forest inventory attributes used for this study were: forest age, diameter at breast height, and species percent. Diameter at breast height is the diameter of the tree trunk at 1.3 m above ground.

Data for elevation and aspect were generated from a DEM provided by the Province of British Columbia. The original elevation model had 25-m^2 grid cells and was created from 1:20,000-scale Terrain Research Information Management data (British Columbia Ministry of Sustainable Resource Management 1996). The data were interpolated using a linear process; the DEM is considered accurate within 10 m. Aspect data was derived from the DEM using the down-slope direction of the maximum rate of change in value from each cell to its neighbours. All data were represented using raster surfaces with a cell size of 200 m^2. This cell size enabled comparison with surfaces created from kernel density estimators, which were used to represent the empirical and randomized infestation intensities (see section 4.2). To generate 200-m^2 cells, nearest neighbour re-sampling was undertaken.

3 Forest Risk Model

In order to implement a spatial randomization procedure that recognizes that the probability of infestation varies over a landscape, we need a means of determining the spatial distribution of forest risk (Bentz et al. 1993; Shore et al. 2000). We used the Shore and Safranyik (1992) risk model which incorporates most of the current knowledge on the biological behaviour of mountain pine beetles and is also the primary model used for forest management in British Columbia. This model (referred in this paper as the risk model) calculates the short-term expectation, or probability, of tree mortality in a stand as a result of infestations (Shore and Safranyik 1992), quantified by loss in stand volume (rather than individual trees).

3.1 Stand Susceptibility

Forest risk is considered to be a function of two components, stand susceptibility and beetle pressure. Stand susceptibility reflects the inherent characteristics of a stand that affect the probability of attack and

damage (Shore and Safranyik 1992). In this context, a stand is defined as a homogenous aggregate of trees. Stand susceptibility (S) ranges from zero to 100, and is calculated as

$$S = PADL \qquad \text{eq 1}$$

where P is the percentage of susceptible pine basal area, A is a pine age factor, D is stand density, and L is a location factor. Each variable used in the calculation of stand susceptibility incorporates a factor important for mountain pine beetle host selection. Originally, susceptibility attributes were determined from look-up tables with discreet categories; however, the model has been updated to allow continuous representation of attributes (Shore and Safranyik 1992; Wulder et al. 2004).

The percentage of susceptible pine basal area (P) is designed to consider tree diameter at breast height and stand composition. P is calculated as

$$P = \frac{(\text{average basal area/ha of pine} \geq 15\,\text{cm dbh})}{(\text{average basal area/ha of all species} \geq 7.5\,\text{cm dbh})} 100 \qquad \text{eq. 2}$$

In Equation 2, the 15-cm diameter at breast height threshold reflects mountain pine beetle preference for mature trees (Hopping and Beal 1948). Under epidemic conditions, smaller trees may be attacked, but fewer beetles will emerge than were required to attack the tree (Safranyik et al. 1974). The lower 7.5-cm threshold is a practical limitation, as smaller trees are not typically included in forest inventories.

Forest age relates directly to a pine tree's ability to resist an attack by mountain pine beetles. Older trees, which are less able to resist attack, are more susceptible to mountain pine beetles. Equations used to calculate the continuous forest-age factor, which ranges from zero to one, are shown in Table 1 (Wulder et al. 2004; Riel, pers. comm.).

Table 1. Age factors used in the Shore and Safranyik forest risk model (1992).

Average pine age (years)	Age factor calculation
≤ 40	0
> 40 to ≤ 80	$0.1 + (0.1((age - 40) / 10)^{1.585})$
> 80 to ≤ 120	1
> 120 to ≤ 520	$1 - (0.05(age - 120) / 20)$
> 520	0

The relationship between tree mortality and stand density reflects factors such as tree vigour and microclimate (wind, light and temperature). Highest pine mortality occurs when stand density is between 250 and 2500 stems per hectare. A density factor ranging from zero to one can be generated using equations found in Table 2 (Wulder et al. 2004; Riel, pers. comm.).

Table 2. Stand density factors used in the Shore and Safranyik forest risk model (1992).

Stand density (stems per ha of trees ≥ 7.5-cm diameter at breast height)	Stand density calculation
< 650	$0.0824(\text{density} / 250)^2$
≥ 650 to < 750	$1 - (0.7(3 - (\text{density} / 250))^5$
≥ 750 to < 1500	1
> 1500	$1 / (0.9 + (0.1\exp(0.4796((\text{density} / 250) - 6))))$

The location factor is an indicator of climate, and is based on latitude, longitude and elevation in British Columbia. It represents the relationships between beetles and climate, whereby higher mortality rates are linked with colder climates. In locations where latitude, longitude, and elevation would suggest that the temperatures are cooler, the value of L is low; locations associated with warmer climates have higher values of L. To determine L, a parameter (Y) is calculated using

$$Y = (24.4\,\text{Longitude}) - (121.9\,\text{Latitude}) - (\text{Elevation (m)}) + (4545.11) \qquad \text{eq. 3}$$

and then the location factor is calculated as shown in Table 3.

Table 3. Location factors used in the Shore and Safranyik forest risk model (1992).

Y	Location factor calculation
> 0	1
≤ 0	$1 / (0.9 + (0.1\exp(-0.8(Y / 250))))$

As with many locations in western Canada, the primary data source available for Morice is the provincial forest inventory data, which does not include basal area or density information. This necessitates modification of the input parameters used to model forest susceptibility. To operationalize the forest risk model, diameter at breast height is often used as a surrogate for density, and basal area is replaced by percentage of pine in a stand (Wulder et al. 2004). We computed susceptibility using the percent of pine in each stand, diameter at breast height (converted to a diameter at breast height factor, see Table 4), and the location and age factors previously outlined.

Table 4. Diameter at breast height (dbh) factors used in the Shore and Safranyik forest risk model (1992).

dbh (cm)	dbh factor
≤ 20	0.1
> 20 and ≤ 22.5	0.6
> 22.5 and ≤ 25	0.8
> 25	1

3.2 Beetle Pressure

The second component of forest risk, beetle pressure (B), is related to both the number and proximity of infested trees. Originally, beetle pressure was calculated for forest stands. Because we represent the forest in raster form here, the version presented has been modified for use with grid cells 200 m^2 in size (Shore, pers. comm.). Beetle pressure is calculated using a two-step procedure. First, the size of the infestation is determined based on the number of attacked trees within a pixel and the number of trees

within 3 km of the pixel (Table 5). Then, based on the infestation size and the proximity of the nearest infestation, beetle pressure is calculated using the equation in Table 6.

Table 5. Grid-based mountain pine beetle infestation size for use in the Shore and Safranyik forest risk model (1992).

Number of infested trees outside pixels and within 3 km	Number of infested trees inside the pixel		
	< 2.5	2.5 to 25	> 25
< 900	Small	Medium	Large
900 to 9,000	Medium	Medium	Large
> 9,000	Large	Large	Large

Table 6. Beetle pressure calculations for the Shore and Safranyik forest risk model (1992).

BP category	Beetle Pressure Calculation
None	0
Small	0.582 – (0.123(nnd/1000))
Intermediate	0.803 – (0.163(nnd/1000))
Large	1.003 – (0.209(nnd/1000))

nnd = nearest neighbour distance; bp = beetle pressure

3.3 Forest Risk

Although susceptibility is a probabilistic measure of forest loss, without information on the beetle population, it provides only a long-term representation of infestation likelihood. Beetle pressure, which represents the state of the current mountain pine beetle population, is necessary for characterizing short-term conditions. Forest risk ranges from zero to one, and is calculated as

$$Risk = 2.74\left(S^{1.77}2.718^{-0.0177*S}\right)\left(B^{2.78}2.718^{-278*B}\right). \qquad \text{eq. 4}$$

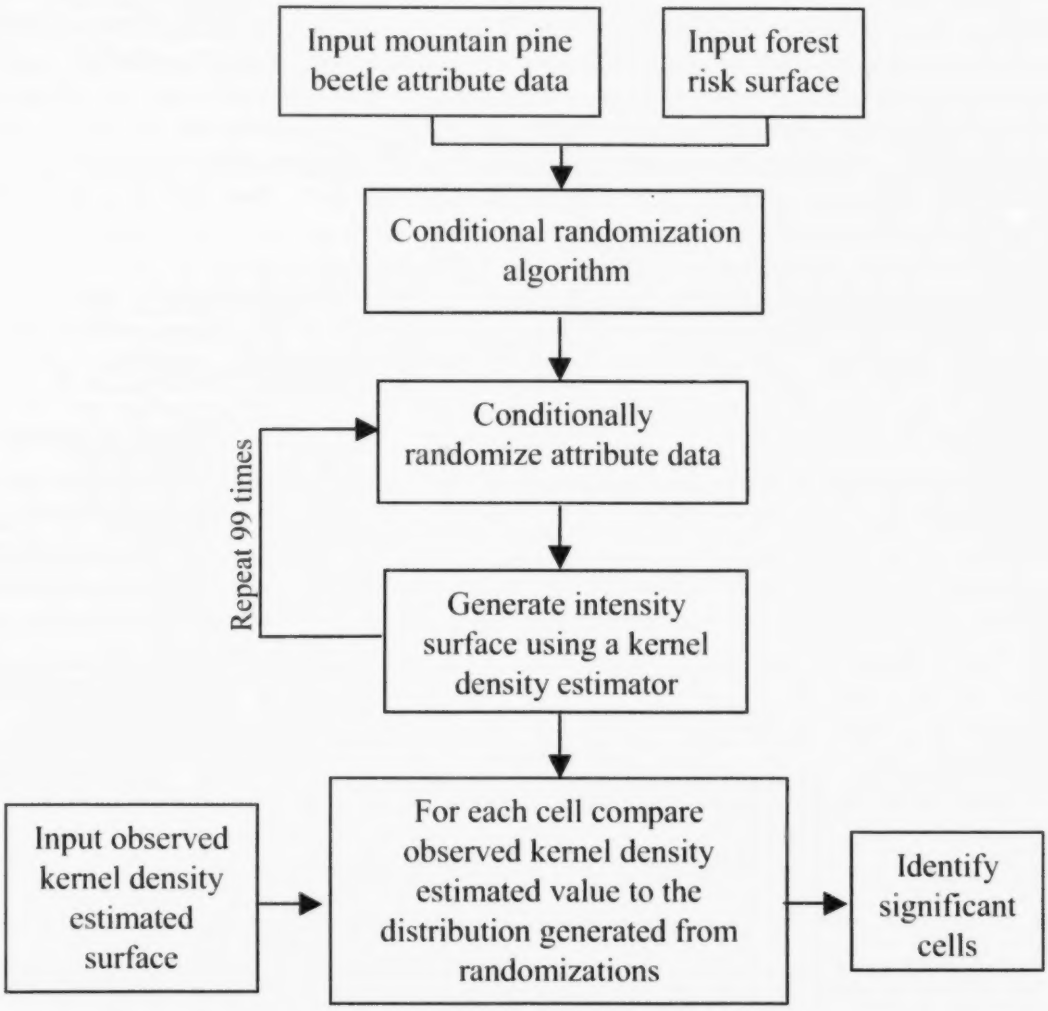
We used the probabilistic representation of risk to condition the spatial randomizations. When data are available for multiple consecutive years, as in this study, risk generated using data in time t can be used to conditionally randomize data in t+1.

4 Identifying and exploring hot spots

An overview of how the randomization procedures were used to identify locations with significantly high infestation intensities is provided in Figure 2. To implement the randomization, observed attribute data and forest risk surfaces were used as input and a randomization algorithm (see Section 4.1) were chosen. Each set of randomized values was then used to generate an infestation intensity surface for the study area using a kernel density estimator (see Section 4.2). This procedure was repeated 99 times. An

intensity surface was also generated in the same fashion for the observed data. All intensity surfaces were represented as raster grids of the same size. For every pixel within the grids, a reference distribution was generated from the values of the test statistics computed for 99 randomizations of the observed data plus the value for the observed data. Statistical significance of the test statistic for the empirical data was determined through comparison with the reference distribution. We labelled statistically significant pixels as hot spots.

Figure 2. An overview of the procedure for identifying statistically significant cells (hot spots).


Input mountain pine beetle attribute data
Input forest risk surface
Conditional randomization algorithm
Conditionally randomize attribute data
Repeat 99 times
Generate intensity surface using a kernel density estimator
Input observed kernel density estimated surface
For each cell compare observed kernel density estimated value to the distribution generated from randomizations
Identify significant cells

4.1 Randomization Algorithms

We considered four algorithms for implementing conditional randomizations (Table 7). These algorithms reflect different ways of conceptualizing the data and mountain pine beetle processes. They can be categorized into theoretical approaches based on whether the randomizations involve individually infested trees or whether they involve infestation clusters. Using an individual tree approach to randomization, we conceptualize the infestation data as single trees on the landscape; using the cluster approach, we consider points to represent groups of infested trees.

In the first algorithm, which uses the individual tree approach, we determine the total number of infested trees from the observed data, and randomly assign individual trees to the landscape using the condition of forest risk. It is possible for more than one tree to be allocated to each pixel; the allocation of each tree is independent of other tree locations. In other words, infested trees are located by an inhomogeneous planar Poisson process, in which the intensity of infested trees is proportional to risk. The difficulty with this algorithm is that it ignores known aggregative processes associated with mountain pine beetles. Mountain pine beetles do not independently select trees to infest; processes of aggregation and dispersion are required to allow successful infestation, with trees near current infestations having a greater likelihood of attack.

Table 7. Overview of conditional spatial randomization algorithms used in the study.

Algorithm	Approach	Allocation Characteristics	Advantages	Disadvantages
1	Individual Tree	Inhomogenous planar Poisson process in which the intensity of infested trees is proportional to risk	Considers infestation clusters as individual trees	Unrealistic in terms of mountain pine beetle processes; frequency distribution of observed data is not reproduced
2	Individual Tree and Cluster	Inhomogenous planar Poisson process in which the presence of infestation clusters is proportional to risk and infestation size grows proportional to risk	Considers infestation occurrence and size proportional to risk	Frequency distributions of observed data is not reproduced; does not adequately reflect mountain pine beetle processes
3	Cluster	Inhomogenous planar Poisson process in which the presence of infestation clusters is proportional to risk	Retains frequency distribution of observed data	Cluster size is not considered
4	Cluster	Inhomogenous planar Poisson process in which the presence and size of infestation clusters is proportional to risk	Retains frequency distribution of observed data; more realistic in terms of beetle behaviour	

The second algorithm combines the individual tree and cluster approaches by first randomly locating the observed number of clusters. Each cluster is given an initial size of one infested tree. Then the remaining observed trees (total numbers of trees – number of clusters) are randomly assigned to cluster sites. For both clusters and trees, the assignment process is random conditioned by forest risk. This algorithm is consistent with mountain pine beetles first selecting locations according to an inhomogenous planar Poisson process, in which the occurrence of an infestation cluster is proportional to risk, and then growing the infestation to a size that is also proportional to risk. If this algorithm is appropriate, it should result in a

distribution of cluster sizes that is similar to that of the observed data. Although several variations of this algorithm were tested, none were able to replicate the observed cluster size distribution.

The third and fourth algorithms are based on clusters and retain the observed frequency distribution of cluster sizes. The third algorithm randomly assigns clusters to locations conditioned on forest risk, without considering the size of the cluster. This is equivalent to an inhomogenous planar Poisson process in which the cluster occurrence is proportional to risk while the cluster size is independent of risk. However, this scenario is inappropriate because risk should reflect the likelihood of loss in stand volume, emphasizing both the presence and magnitude of mountain pine beetle infestations.

The fourth algorithm assigns clusters to locations based on an inhomogenous planar Poisson process in which both cluster occurrence and size are proportional to risk. To achieve this, the clusters are divided into n ordered groups (g_1, ..., gi, ..., gn) on the basis of size. Similarly, the locations are divided into n ordered groups (r_1, ..., ri, ..., rn) on the basis of magnitude of risk. Then the clusters in group gi are randomly assigned to locations in the corresponding group, ri. Although this algorithm has a stochastic component, it generally forces the largest clusters to be in the highest risk locations. The greater the number of groups, the less the stochastic component is the assignment process. When there is only one group, this algorithm is equivalent to Algorithm 3. In this study, n was set equal to three based on natural breaks in the mountain pine beetle-attribute data. Algorithm 4 reflects biological theory on the mountain pine beetle, which suggests that higher risk areas are likely to have more and larger clusters than are areas with less risk. Additionally, it has the benefit of retaining the empirical distribution of cluster sizes. Therefore, we selected Algorithm 4 for operationalizing the conditional randomizations.

4.2 Test Statistic

The test statistic is the estimated value of the intensity of mountain pine beetle infestation at a given location for the empirical data. These values are obtained using a non-parametric kernel density estimator (Silverman 1986; Bailey and Gatrell 1995) that enables the observed, aerial point data to be converted to a raster surface format in which the pixel values represent the number of infested trees per unit area. The intensity $\lambda(z)$ at a particular location z in a study area A can be estimated by

$$\hat{\lambda}_\tau(z) = \frac{1}{p_\tau(z)} \left\{ \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{(z - z_i)}{\tau} \right) y_i \right\} \quad z \in A \qquad \text{eq. 5}$$

where τ is the radius of a disk centered on z, k( ) is the kernel or a probability density function that is symmetric around about the origin; zi (i = 1, ..., n) is the location of n observed events; and yi is the attribute value at zi. The term $p_\tau(z) = \int_A k[(z - u)/\tau] du$ is an edge correction equivalent to the volume under the scaled kernel centred on z, which lies inside of A (Diggle 1985).

There are three issues that are commonly considered when working with kernel estimators: the type of kernel k( ), the size of disk radius τ, and edge effects. The kernel determines how events within the disk will be weighted. Here we use a kernel with a quartic distribution function given by

$$\hat{\lambda}_\tau(z) = \frac{1}{p_\tau(z)} \left\{ \sum_{h_i \le \tau} \frac{3}{\pi \tau^2} \left(1 - \frac{h_i^2}{\tau^2} \right)^2 y_i \right\} \quad z \in A \qquad \text{eq. 6}$$

where hi is the distance between the point z and the observed event location zi. Although kernel type may be theoretically important, it does not have a large impact on kernel output.

Kernel estimators are much more sensitive to τ, which controls data smoothing. As τ increases, so does the amount of data smoothing (Kelsall and Diggle 1995); if τ is too large, data variability will be lost, whereas if it is too small, trends in data will not be visible. If the study area is a unit square, a good starting point for determining an appropriate value for τ is $0.7n^{-0.2}$. However, kernels often need to be calculated for several values of τ and kernel outputs need to be compared. In this study, we compared several different kernel sizes, and chose to use a 2-km disc radius, optimizing tradeoffs between detail and representations of infestation trends.

Kernel estimators may also be affected by edge effects. In this study, edge effects do not have a large impact, as the area we are studying is large relative to τ, and so the edge correction term in equation 5 was not implemented. An additional issue that arises in representing kernel estimator values in a raster format is the definition of surface cell size. We used a 200-m^2 grid cell, as the data points represented circular areas with a maximum diameter of 200 m.

The empirical data was randomized 99 times; each time, the associated intensity surface was computed using the kernel density estimator. The test statistic, which is the value for a given pixel in the intensity surface generated from the empirical data, is evaluated by examining its position within the reference distribution of intensity values composed of its value plus the 99 values for the corresponding pixel in the surfaces generated using the randomized data. Thus, the smallest significance level that can be used for statistical testing is 0.01. If the observed value exceeded all 99 values from the randomizations, the observed value was considered significantly different from chance at the 0.01 significance level and the pixel was labelled a hot spot.

This procedure was repeated for each annual set of observed data, from 1996 to 2002. The local nature of the test statistic allows the spatial variation in hot spots to be mapped and visualized.

4.3 Exploring hot spots

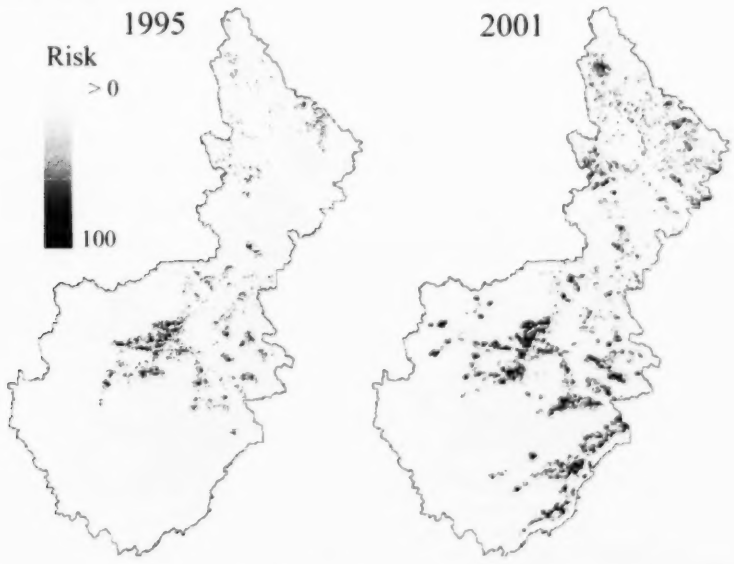


Figure 3. The probability of forest risk calculated for the Morice Timber Supply Area for 1995 and 2001.

Selective landscape characteristics underlying hot spots were compared with those for the entire study area. For each sub-area of Morice (north, middle, and south), the relative frequency distributions for pine age, percent pine, elevation, and aspect underlying hot spots were compared to those for all pine cells (i.e., all cells that could host infestations). Owing to the large number of possible comparisons, only selected results representative of the general trends that were observed are presented in the following section.

5 Results and discussion

5.1 Conditional Randomization

The forest risk surfaces calculated for 1995 and 2001, which were used to condition the randomizations of the 1996 and 2002 empirical data, respectively, are shown in Figure 3. The probability of forest risk ranges from zero to 100. For risk to equal zero, susceptibility must also be zero. Because temporal changes to susceptibility are subtle, annual variations in risk are primarily the result of changes in beetle location.

Table 8. The percentage of pixels with risk greater than 0 that were hot spots in Morice and its sub-areas, from 1996 to 2002.

		1996	1997	1998	1999	2000	2001	2002
Sub-area	North	26.95	22.91	8.17	19.02	17.89	25.16	6.32
	Middle	4.18	12.66	8.43	7.32	6.84	13.35	3.30
	South	0.00	0.00	2.69	6.13	11.16	9.67	45.08
Total	Morice	10.43	13.26	7.17	10.72	11.18	16.28	12.81

Between 1996 and 2002, on average, 11.69% of the locations at risk of infestation in the study area had intensity values that were significantly greater than those expected by chance (see Table 8). Although the annual extent of infestation is relatively stable for the entire study area—ranging from 7.17% to 16.28%—there is more variability in the sub-areas. From the locations of hot spots for 1996 and 2002 shown in Figure 4, it is apparent that hot spots cluster spatially. Collectively, these results indicate that hot spots occur most often where the infestation is most active. For example, in 1996 the most extensive incidence of hot spots was in the north, but by 2002 the majority of them were found in the south. This reflects trends in the intensity of mountain pine beetle activity in Morice. High-magnitude infestations initially occurred in the north and middle sub-areas, but by 2002 the mountain pine beetle population was largest in the south.

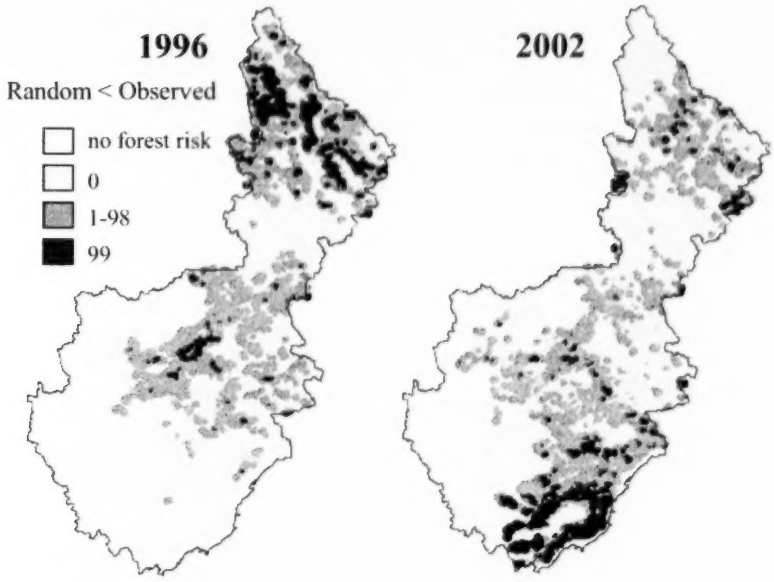


Figure 4. Number of conditional randomizations less than observed value. Black indicates statistically significant (α=0.01) locations (hot spots)

5.2 Comparisons with landscape characteristics

Frequency distribution of pine age associated with hot spots in a sub-area becomes more similar to age-frequency distribution of all pine as infestations progress through time (Figure 5). Most commonly, the age of forests at hot spot locations are 180 years in the north sub-area, 140 years in the middle sub-area, and older than 180 years in the south sub-area. Why the middle sub-area is different is unclear, but may be related to the area's lower proportion of trees that are 180 years old.

The percentage of pine associated with hot spots varies markedly for each sub-area (Figure 6). In the north sub-area, hot spots tend to be associated with low percentage pine (30%); in the middle sub-area, they are more associated with high percentage pine (60% to 70%); in the south sub-area, moderate percentage pine (30% to 50%) is dominant. In the middle and south sub-areas, as the infestation persists, the percent pine frequency distributions of hot spots become more similar to the general percent pine frequency distributions.

When hot spots are compared in terms of elevation, a larger proportion of cells is associated with lower elevations than for all locations (Figure 7). In the north sub-area, there is no dominant elevation consistently related to hot spots. In the middle sub-area, hot spots are consistently strongly associated with elevations of 800 m, whereas in the south sub-area, elevations of 800 to 900 m prevail. There also seems to be a trend whereby hot spots are associated with lower elevations as latitudes become more northerly. Although additional data on climate is necessary to investigate further elevation trends, it appears that as latitudes increase, temperatures warm enough to support large mountain pine beetle populations occur at lower elevations.

Hot spots show meaningful relationships with aspect only for the south sub-area (Figure 8). In general, south- and southeast-facing slopes are more related to hot spots, and through time, the frequency distribution associated with hot spots becomes more similar to the aspect frequency distribution for all pine locations.

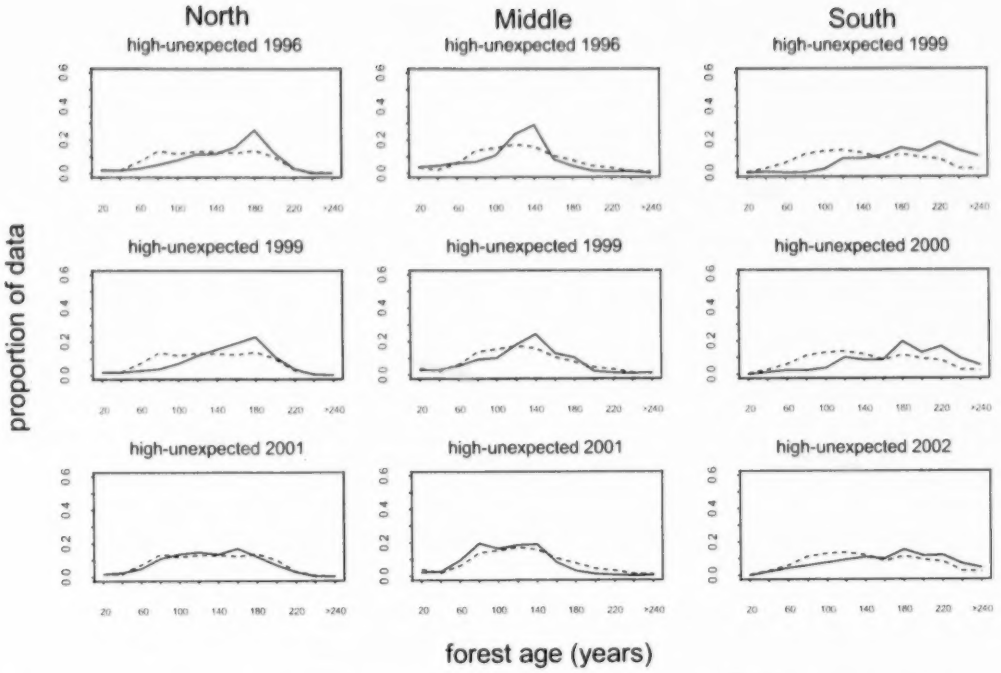


Figure 5. Distribution of forest age by sub-area in Morice Timber Supply Area: hot spots = solid line; all locations with pine in sub-area = dotted line.

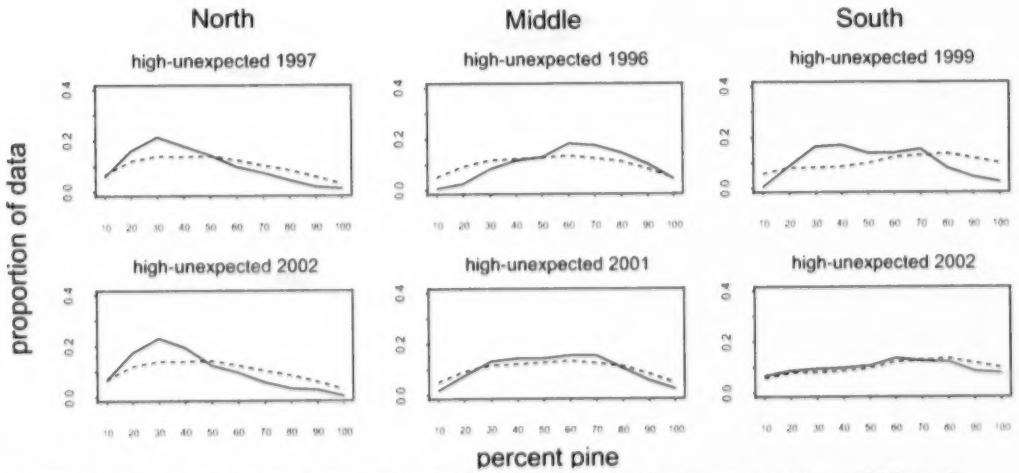


Figure 6. Distribution of percent pine by sub-area in the Morice Timber Supply Area: hot spots = solid line; all locations with pine in sub-area = dotted line.

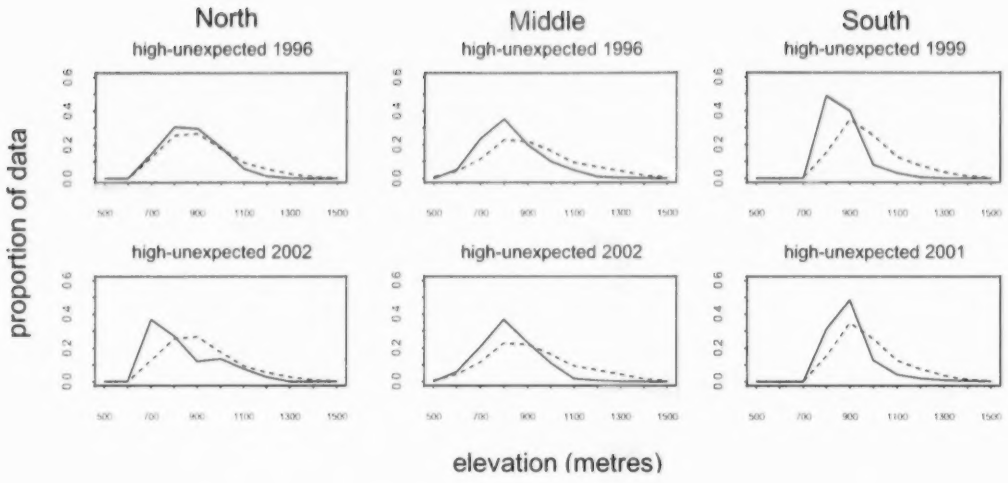


Figure 7. Distribution of elevation by sub-area in the Morice Timber Supply Area: hot spots = solid line; all locations with pine in sub-area = dotted line.

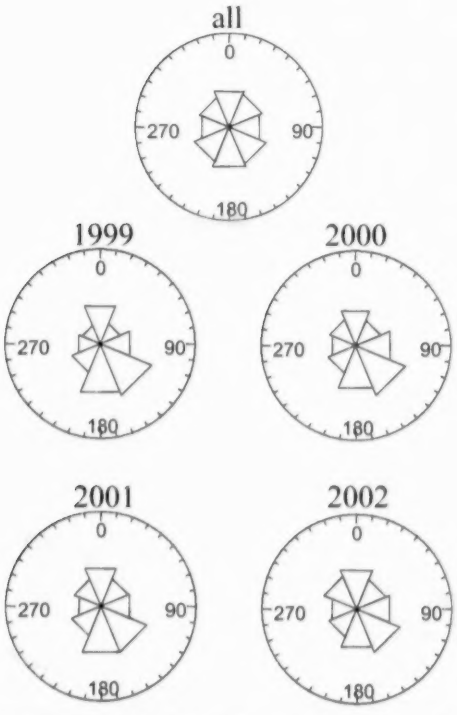


Figure 8. Distribution of aspect in the South sub-area of the Morice Timber Supply Area: all locations with pine and hot spots by year.

6 Conclusions

Because little is currently known about the factors that influence spatial patterns of mountain pine beetle infestations, we used a randomization procedure to undertake an exploratory analysis of observed patterns to evaluate if the patterns could have arisen by chance in the study area. As landscape conditions vary over our study area, we felt it was inappropriate to consider chance conditions in terms of complete spatial randomness; instead, we used an existing model of forest risk—which incorporates both forest conditions and beetle occurrence—to determine the probability of infestation occurring at a given location. By representing infestation in terms of intensity surfaces represented in raster format, we were able to evaluate each location (pixel) in our study area.

Our analysis revealed that, for the years between 1996 and 2002, the percentage of locations in the Morice study area that could be considered to have infestation values significantly greater than were expected by chance ($\alpha = 0.01$) ranged from 7.17% to 16.28%, although more extensive coverage was apparent when the results were stratified by sub-area. Trends in hot-spot locations and extent reflected those in the intensity of mountain pine beetle activity in Morice, with the most extensive incidence of hot spots occurring when and where infestation was most active. This suggests that, although there is a random component to mountain pine beetle activity, this component becomes less as infestation levels increase. The presence of a systematic influence on the patterns is also indicated by our finding that hot spots cluster spatially in all years. In order to explore this further, we examined whether selected forest and topographic conditions associated with hot spots were distinctive in some way.

In terms of forest conditions, we found that hot spots were distinctive in terms of both age of pine and percentage pine. In particular, hot spots were associated most often with low values of the latter. This suggests that further research on the roles these factors play in mountain pine beetle biology is warranted. However, the distinctive features we observed were neither consistent across the three sub-areas of our study area nor across the time period examined. This suggests that the presence of additional local influences, as well as possible differences in the way factors considered interact over space and time.

In terms of topography, although we found that likelihood of hot spots increased more than was expected at low elevations in all sub-areas, this tendency increases as latitudes become more northerly. Such elevations are associated with warmer temperatures that would result in greater susceptibility to mountain pine beetle infestation. However, aspect was found to be influential only in the south sub-area, where hot spots are associated with south-facing and southeast-facing slopes, particularly during the initial phase of infestation. This finding is important because aspect is not often considered as an influence on mountain beetle infestations. Taken together, the results relating to topography point to the influence of climatic factors, especially temperature. Clearly, there is need for further investigation of how patterns of infestations relate to spatial variability in climate over the study area.

In aggregate, there is a general tendency in all sub-areas for hot spots to become less distinctive in terms of both forest and topographic features over the duration of the infestation. This implies that the impacts of the features examined become increasingly location specific, that other, as yet unidentified, factors become important, or that both possibilities may occur.

It is important to acknowledge that our results depend on several implicit assumptions being satisfied. The first of these is that the empirical data are reliable. Because of the nature of aerial GPS surveys, we recognize that there will be some degree of error involving both the locations and sizes of the clusters. However, such error is ameliorated by representing the data as intensity surfaces generated by kernel density estimators that have the effect of smoothing the data. Similarly, we recognize that both the forest inventory data and the digital elevation model are also subject to error. We also assume that the forest risk model (Shore and Safranyik 1992) used to condition our randomizations is accurate. Although there is no way of testing this directly, it is the most suitable model currently available for this work. Finally, as we show, there is more than one way of randomizing the data even when forest risk is known. Clearly, further work is necessary to determine the sensitivity of our results to all of these influences.

Acknowledgements

This project was funded by the Government of Canada through the Mountain Pine Beetle Initiative, a six-year, $40-million program administered by Natural Resources Canada, Canadian Forest Service. Publication does not necessarily signify that the contents of this report reflect the views or policies of Natural Resources Canada, Canadian Forest Service. The authors are grateful for the constructive comments of two anonymous reviewers that helped enhance our presentation.

Contacts

Trisalyn Nelson
Department of Geography
University of Victoria
Victoria, British Columbia V8W 3P5
Canada
Telephone: 250-721-7327; Fax: 250-721-6216; Email: tnelson@mail.geog.uvic.ca

Barry Boots
Department of Geography and Environmental Studies
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5
Canada
E-mail: bboots@wlu.ca

References

Bailey, T.; Gatrell, A. 1995. *Interactive Spatial Data Analysis*. Longman Group Limited, Essex.

Barclay, H.; Safranyik, L.; Linton, D. 1998. Trapping mountain pine beetles *Dendroctonus ponderosae* (Coleoptera: Scolytidae) using pheromone-baited traps: effects of trapping distance. Journal of the Entomology Society of British Columbia 95:25–31.

Bentz, B.; Amman, G.; Logan, L. 1993. A critical assessment of risk classification systems for the mountain pine beetle. Forest Ecology and Management 61:349–366.

Besag, J.; Newell, J. 1991. The detection of clusters in rare diseases. Journal of the Royal Statistical Society 154:143–155.

Boots, B. 2002. Local measures of spatial association. Ecoscience 9:168–176.

Boots, B.N.; Getis, A. 1988. *Point Pattern Analysis*. Scientific Geography Series, Volume 8. Sage Publications, Newbury Park, California .

Conner, E,; Simberloff, D. 1986. Competition, scientific method, and null models in ecology. American Scientist 74:155–162.

Davis, J.H.; Howe, R.W.; Gregory, J.D. 2000. A multi-scale spatial analysis method for point data. Landscape Ecology 15:99–114.

Diggle, P. 1985. A kernel method for smoothing point process data. *Applied Statistics* 34: 138–147.

Edgington, E.S. 1995. *Randomization Tests*. Third edition. Marcel Dekker Inc., New York.

Fortin, M.J.; Jacquez, G. 2000. Randomization tests and spatially autocorrelated data. Bulletin of Ecological Society of America 81:201–205.

Fotheringham, A. 1997. Trends in quantitative methods I: stressing the local. Progress in Human Geography 21:88–96.

Fotheringham, S.; Brunsdon, C. 1999. Local forms of spatial analysis. Geographical Analysis 31:340–358.

Getis, A.; Boots, B. 1978. Models of Spatial Processes: An Approach to the Study of Point, Line and Area Patterns. Cambridge University Press, Cambridge.

Hopping, G.R.; Beall, G. 1948. The relation of diameter of lodgepole pine to incidence of attack by the bark beetle *Dendroctonus monticolae* Hopkins. Forestry Chronicle 24:141–145.

Kelsall, J.; Diggle, P. 1995. Non-parametric estimation of spatial variation in relative risk. Statistics in Medicine 14:2335–2342.

Kulldorff, M.; Tango, T.; Park, P.J. 2003. Power comparisons for disease clustering tests. Computational Statistics & Data Analysis 42:665–684.

Leckie, D.; Gillis, M. 1995. Forest inventory in Canada with emphasis on map production. Forestry Chronicle 71: 74–88.

Legendre, P.; Dale, M.R.T.; Fortin, M.J.; Gurevitch, J.; Hohn, M.; Myers, D. 2002. The consequence of spatial structure for the design and analysis of ecological field surveys. Ecography 25:601–615.

Legendre, P.; Fortin, M.J. 1989. Spatial pattern and ecological analysis. Vegetation 80:107–138.

Mitchell, R.; Preisler, H. 1991. Analysis of spatial patterns of lodgepole pine attacked by outbreak populations of the mountain pine beetle. Forest Science 37:1390–1408.

O'Sullivan, D.; Unwin, D.J. 2003. *Geographic Information Analysis*. John Wiley & Sons Inc., Hoboken, N.J.

Parminter, J. 1998. Natural Disturbance Ecology. *Pages* 3–35 *in:* Voller, J.; Harrison, S. (eds) *Conservation Biology Principles for Forested Landscapes*. UBC Press, Vancouver.

British Columbia Ministry of Sustainable Resource Management. 1996. Gridded DEM specifications. Victoria, British Columbia.

Safranyik, L.; Shrimpton, D.; Whitney, H. 1974. Management of lodgepole pine to reduce losses from the mountain pine beetle. Government of Canada. Department of the Environment. Canadian Forest Service, Pacific Forest Research Centre, Victoria, BC. Forestry Technical Report 1.

Shore, T.; Safranyik, L. 1992. Susceptibility and risk rating systems for the mountain pine beetle in lodgepole pine stands. Forestry Canada, Pacific Forestry Centre, Victoria, BC. Information Report BC-X-336. 12 p.

Shore, T.; Safranyik, L.; Lemieux, J. 2000. Susceptibility of lodgepole pine strands to the mountain pine beetle: testing of a rating system. Canadian Journal of Forest Research 30:44–49.

Silverman, B. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman Hall, New York

Stine, P.; Hunsaker, C. 2001. An introduction to uncertainty issues for spatial data used in ecological applications. *Pages* 91–107 *in:* Hunsaker G, Goodchild M, Friedl M, Case T (eds) Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications. Springer, New York.

Westfall, J. 2004. *2004 Summary of Forest Health Conditions in British Columbia*. British Columbia Ministry of Forests, Forest Practices Branch. 43 p.

Wulder, M.A.;, Seeman, D.; Dymond, C.; Shore, T.; Riel, B. 2004. Arc/Info Macro Language (AML) scripts for mapping susceptibility and risk of volume losses to mountain pine beetle in British Columbia. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoira, BC. Technology Transfer Note 33. 4 p.

This publication is funded by the Government of Canada through the Mountain Pine Beetle Initiative, a program administered by Natural Resources Canada, Canadian Forest Service (web site: mpb.cfs.nrcan.gc.ca).

Contact:

For more information on the Canadian Forest Service, visit our web site at:
www.nrcan.gc.ca/cfs-scf

or contact the Pacific Forestry Centre
506 West Burnside Road
Victoria, BC V8Z 1M5
Tel: (250) 363-0600 Fax: (250) 363-0775
www.pfc.cfs.nrcan.gc.ca

To order publications on-line, visit the Canadian Forest Service Bookstore at:
bookstore.cfs.nrcan.gc.ca